基于演化事件探测的学科领域科研社群演化特征研究*
——以图书馆学情报学为例

■ 李纲[1] 唐晶[1] 毛进[1] 田云裴[1] 张斌[2]
[1] 武汉大学信息资源研究中心 武汉 430072 [2]南京大学信息管理学院 南京 210023

摘 要：[目的/意义]科研社群作为当代科学研究中重要知识群体，研究其在学科发展过程中的演化特征对于探索领域发展规律、促进知识创新等方面具有重要意义。[方法/过程]为探究科研社群的动态演化特征，以图书馆学情报学领域为例，从演化事件探测的角度出发，采用 Leiden 算法对科研社群进行划分，并构建科研社群演化路径与演化树；在此基础上识别科研社群演化事件，从科研社群演化整体分析、科研社群演化路径及演化树特征分析、科研社群演化事件统计特征分析等 3 个方面来揭示科研社群的演化模式和演化特征。[结果/结论]研究表明，科研社群规模呈蓬勃发展趋势，科研社群演化树呈现两种演化模式，增长类演化事件大多发生于大型科研社群且发文量较高，新生和消亡演化事件均发生于小型科研社群且发文量较高，合并、部分合并、分裂、衰减等演化事件的平均社群规模较小且其发文量偏低，进一步证明科研社群之间合作交流趋向频繁，科研社群演化日趋复杂。

关键词：科研社群 Leiden 算法 演化事件 演化特征
分类号：G251
DOI：10.13266/j.issn.0252-3116.2021.17.008

1 引言

随着第四科研范式的逐渐形成和跨学科研究的兴起，科研合作已经成为研究人员解决复杂研究问题的有效途径。科研团队作为科研合作的基本载体，是科学技术研究的主体，由优势互补、愿意为共同的科研目的、科研目标和工作方法而相互协作配合承担责任的科研人员组成的物理群体[1]。但因研究问题日益复杂化与系统化，出现了以科研协作关系为主导的虚拟科研社群[2]。本文中科研社群是指研究人员因合著关系形成的研究群体[3]，识别科研社群并深入探究科研社群的演化特征对于探索领域发展规律、制定科技发展政策、培养创新人才等方面均有重要学术价值与战略意义。

早期的科研社群相关研究主要从静态视角出发，对于静态科研合著网络识别科研社群，并运用信息计量学、社会网络分析等方法对科研社群的数量特征、拓扑结构等进行分析[4-5]。然而实际科研社群会随时间动态变化，如科研社群中有成员加入或退出，成员之间展开新的合作或停止原有合作，传统的静态方法无法深入分析科研社群中的持续演化过程，因此从动态视角探究科研社群演化成为了研究者们的关注重点[6]。

当前研究主要通过定义、识别演化事件并分析各时间窗口的网络属性来研究科研社群演化，但大多数研究对于演化事件的定义并不全面，且未能深入揭示科研社群的演化特征及其规律。基于此，本文以图书馆学情报学(Library and Information Science，LIS)领域为例，采用 Leiden 算法划分科研社群并识别 12 种演化事件，通过分析科研社群整体演化情况、科研社群演化路径及演化树特征、科研社群演化事件统计特征等来揭示科研社群的演化特征及模式。

2 相关研究

当前大多数研究均基于不同时间窗口将网络转换

* 本文系国家自然科学基金创新研究群体项目“信息资源管理”(项目编号:71921002)和 ISTIC-Taylor&Francis Group 学术前沿观察联合实验室开放基金(项目编号:IT1910)研究成果之一。
作者简介：李纲(ORCID:0000-0001-5573-6400)，教授，博士生导师；唐晶(ORCID:0000-0002-1211-5812)，硕士研究生；毛进(ORCID:0000-0001-9572-6709)，副教授，硕士生导师，通讯作者，E-mail:danveno@163.com；田云裴(ORCID:0000-0003-0084-6711)，硕士研究生；张斌(ORCID:0000-0002-5591-7874)，副教授，硕士生导师。
收稿日期:2021-03-14 修回日期:2021-06-07 本文起止页码:79-90 本文责任编辑:徐健

成静态图来探究科研社群演化。这些研究大多分为两个步骤,即科研社群识别和科研社群演化分析。

2.1 科研社群识别

科研社群是由研究人员及其合著关系形成的,前人大多基于主题或拓扑结构来对其进行识别[3]。从主题的角度看,研究人员因存在共同研究主题偏好而形成科研社群,如陈文杰基于科研合作网络的特性提出一种融合作者主题特征的社群发现算法[7],S. Jung 等将具有共同主题偏好的作者聚合成“共同兴趣小组”,并在此基础上探究其主题的演化[8]。J. Mao 等则认为科研社群同时兼具拓扑凝聚性和主题凝聚性,并提出一种主题科研社群识别方法,进而研究了主题科研社群的特征[3]。

从拓扑结构的角度来看,科研合著网络因研究人员之间的合作关系强度而呈现出明显的科研社群结构。研究者大多使用复杂网络中的社区发现算法来对科研社群进行识别,如 Infomap[9]、Label Propagation[10]和 Louvain[11]算法等。Louvain 算法是一种基于模块度的图聚类算法,包含节点局部移动和网络聚合两步迭代设计。在局部移动阶段,算法将各个节点移动到其邻居节点所在的社区,衡量其模块度收益直至最大。在聚合阶段,算法将根据在节点局部移动阶段得到的每个社区折叠成一个节点形成聚合网络,并计算该网络中节点之间的连边权重。最后重复两个阶段,直到整个网络的模块度收益最大化。相较于其他社区发现算法,Louvain 算法具有运行效率高、划分效果好等优点[12],成了研究人员进行科研社群识别的首选算法。如余厚强[13]、邹本涛[14]等采用 Louvain 算法来识别人工智能领域的科研团队,并进一步探究该领域领军团队的特征以及高产团队演化的特征和规律。江文华等[15]在构建图书馆领域核心作者研究兴趣相似性网络的基础上,利用 Louvain 算法对研究者进行社群划分并计算各社群研究兴趣相似性。

然而在 Louvain 算法中,若节点为两个社群的连接点,当将其从原有社群移动到具有模块化最大增益的新社群时,原有社群将会出现内部连接不良甚至断开的现象。针对这一问题,V. Traag 等人提出 Leiden 算法,在 Louvain 算法的基础上使用快速局部移动过程来初步划分社区,提高算法运行速度,并增加分区优化步骤对初步划分社区进行精炼,在保留 Louvain 算法的优越性基础上,进一步提升社群划分效果[16]。

2.2 科研社群演化分析

关于科研社群演化分析的研究大致可分为两方面:一是通过探究科研社群结构变化来反映科研社群的演化特征[17-18],如邹本涛等通过计算节点数、边数、网络密度和平均集聚系数 4 项网络指标的极值分布,从微观和宏观的视角探讨人工智能领域高产科研团队演化特征规律[14]。王曰芬等根据平均集聚系数、平均路径长度等指标变化分析新能源领域的科研合作演变特征[19]。二是通过识别关键演化事件来探究社群演化规律,如 S. Asur 等将社群演化归类为延续、合并、分裂、新生和消亡等 5 项演化事件并从网络规模和网络成员等方面探讨社群演化特征[20],M. Takaffoli 等在 S. Asur 等的研究基础上增加了增长和衰减两种演化事件,并在动态网络中跟踪社群演变[21],G. Palla 等则在识别演化事件的基础上分析演化事件与社群规模大小、社群存在时长之间的相关性[22]。

然而当前关于科研社群演化特征的研究还存在局限性:一方面基于网络拓扑结构变化的科研社群演化跟踪仅能反映科研社群演化整体面貌,另一方面前人研究将科研社群演化事件定义得过于简洁[23],无法反映社群演化的真实情境,一定程度上限制了对科研社群演化特征的深入分析。此外,当前研究在揭示科研社群演化路径并归纳科研社群演化模式方面还存在不足,探究科研社群演化特征的指标还停留在科研社群基本数量特征层面[24],如社群年龄、社群规模大小等,未结合科研产出这一维度对科研社群演化特征进行挖掘。

K. Mohammadmosaferi 等考虑到社群在演化过程中还可能出现部分演化的现象,如两个社群的部分成员合并成了新的社群,因此定义 12 种社群演化事件,更深入地揭示社群演化过程[23]。基于此,本文提出一种探究学科领域中科研社群演化特征的方法,以图书馆学情报学领域为例,使用 Leiden 算法进行科研社群划分,并识别 K. Mohammadmosaferi 等定义的 12 种科研社群演化事件,通过分析科研社群网络特征、演化路径及演化事件来揭示科研社群的内在演化特征及模式,以期更完整地跟踪社群演化过程,推动科研社群发现的理论研究和实践应用。

3 研究设计

本文提出一种从演化事件探测角度出发,探究学科领域中科研社群演化特征的方法,基于更全面的演化事件对科研社群演化过程进行分析,并以图书馆学情报学领域为例展开实证研究,如图 1 所示:

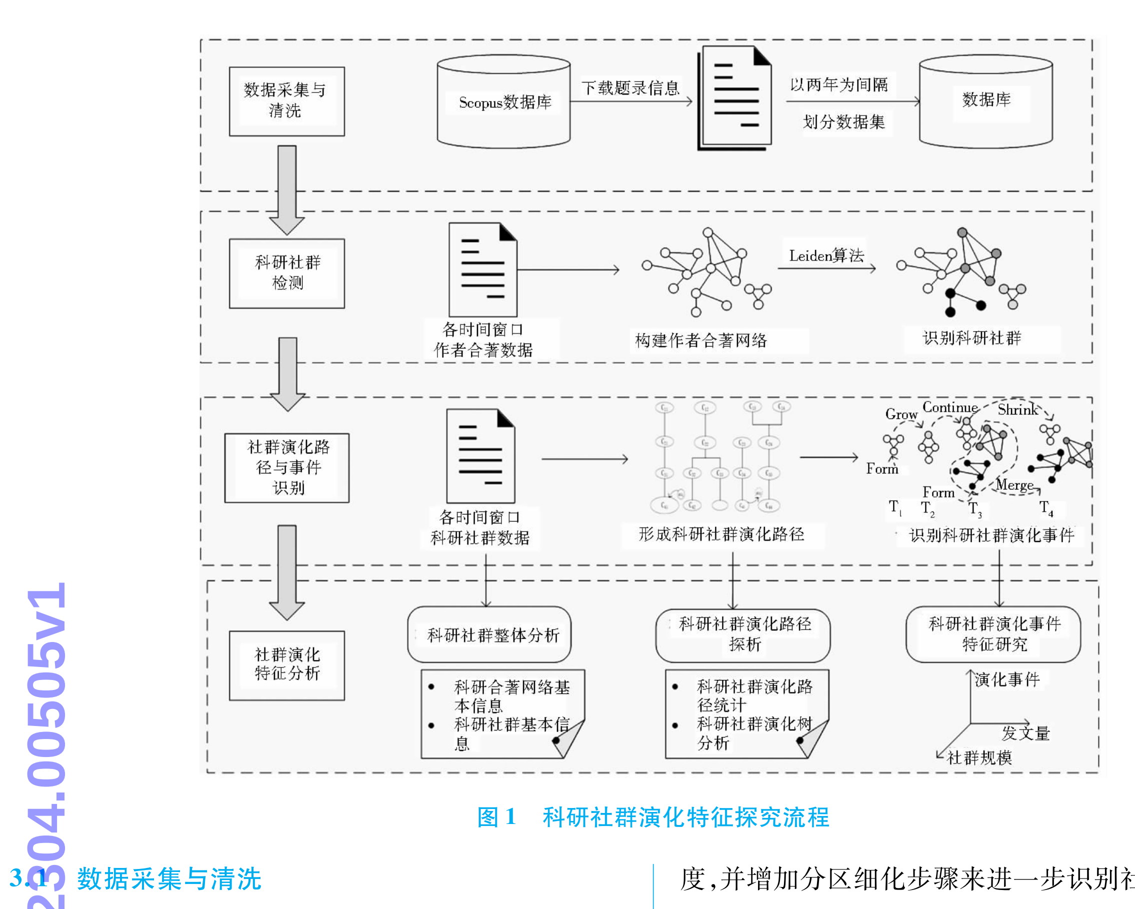


图1　科研社群演化特征探究流程

3.1　数据采集与清洗

在数据采集与清洗阶段，因Scopus数据库为每一位作者分配了唯一的作者标识，可有效解决作者人名歧义问题，本文选取Scopus数据库为数据源，采集图书馆学情报学领域CiteScore较高的前100个期刊中发表于2000－2019年的所有文献，共计78 483篇，经数据清洗后得到文献70 015篇，其中包含作者106 100位。

3.2　科研社群检测

科研社群识别是分析科研社群演化的重要环节。本文基于科研合著网络的拓扑结构，采用社区发现算法识别科研社群。为更好地探究科研社群的演化态势，本文根据发文量过滤掉发文频次小于2的低频作者，并依据2年为一个时间窗口的划分标准，将2000年－2019年图书馆学情报学领域的科研合著情况划分为10个阶段。针对各个时间窗口的合著数据，以作者为网络节点，以合著关系作为边，形成10个科研合著网络。

针对Louvain算法可能会识别内部连接不良甚至断开的社群这一问题，Leiden算法在其基础上作出改进，主要包括节点局部移动、分区细化及基于细化分区的网络聚合3个步骤。该算法在Louvain算法的基础上采用快速局部移动的方法来加快第一阶段的运行速度，并增加分区细化步骤来进一步识别社群，保证识别的社群质量同时进一步提升了算法效率[25]。因此，为更进一步探究科研社群演化过程及规律，本文使用python中igraph包下的Leiden算法，将目标函数设置为模块度(modularity)，将作者之间的合作次数作为边权重，分别对10个科研合著网络进行科研社群识别。同时，因三元闭包可以作为知识网络分析的基本单元[26]，本文选取社群成员大于2的科研社群进行分析，并以“时间窗口，社群编号”格式对社群命名。

最后，将10个科研合著网络数据及社群划分信息输入Cytoscape[27]，计算各阶段合著网络及科研社群的网络特征并进行可视化。

3.3　科研社群演化事件定义与演化路径构建

识别科研社群后，抽取科研社群演化路径与演化树并识别科研社群演化事件是探究科研社群演化的关键。图2展示了一种可能的科研社群演化情景，au表示科研社群成员，箭头方向表示该成员加入或退出该社群。本文将C_n表示为时间窗口n下所有社群的集合：

$$C_n=\{C_{(n,1)},C_{(n,2)},\cdots\cdots,C_{(n,m)},\},T=n$$

公式(1)

公式(1)中T＝n表示时间窗口n，$C_{(n,m)}$表示该时间窗口下某一社群。如图2中社群$C_{(1,1)}$、$C_{(1,2)}$、$C_{(1,3)}$、

$C_{(1,4)}$、$C_{(1,5)}$组成了时间窗口 1 下社群集合 C_1。

在科研社群中,演化路径 P 是相邻时间窗口中存在共同成员关系的科研社群序列,展示出科研社群从新生到消亡过程中社群成员的流向:

$$P_x = [C_{(n,y)}, \cdots\cdots, C_{(m-1,i)}, C_{(m,j)},], n < m-1 < m \qquad \text{公式(2)}$$

公式(2)中 $C_{(m-1,i)}$ 与 $C_{(m,j)}$ 的交集不为空。如图 2 中[$C_{(2,3)}$, $C_{(3,4)}$, $C_{(4,3)}$]、[$C_{(2,3)}$, $C_{(3,4)}$, $C_{(4,4)}$]等即为演化路径。

演化树 Tree 则是指含有相同新生社群的演化路径集合,可直观的表示科研社群演化历程:

$$Tree = \{P_1, \cdots\cdots, P_j\} \qquad \text{公式(3)}$$

公式(3)中所有演化路径 P 均含有相同新生社群。如图 2 中集合{[$C_{(2,3)}$, $C_{(3,4)}$, $C_{(4,3)}$]、[$C_{(2,3)}$, $C_{(3,4)}$, $C_{(4,4)}$]}具有相同新生社群 $C_{(2,3)}$,因此该两条演化路径集合即为一棵演化树。本文在获得各时间窗口的科研社群数据的基础上,匹配相邻时间窗口中各个科研社群的共同成员,并按时序串联形成科研社群的演化路径,最后将含有相同新生社群的演化路径聚合,形成演化树。

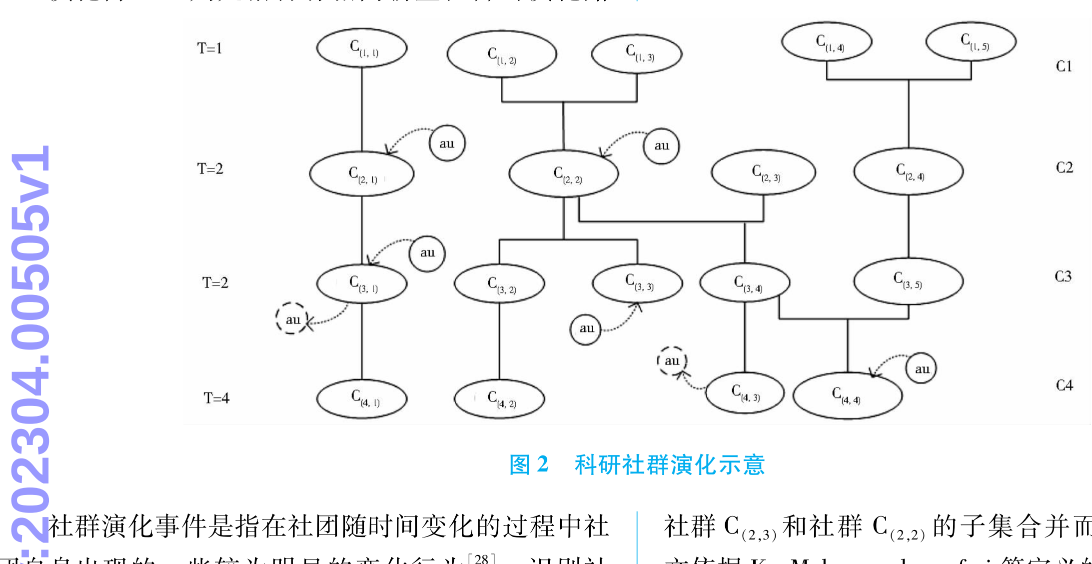


图 2 科研社群演化示意

社群演化事件是指在社团随时间变化的过程中社团自身出现的一些较为明显的变化行为[28]。识别社群演化事件对于分析科研社群的演化有着重要作用。前人研究大多将演化事件分为新生、增长、分裂、合并、衰减、消亡、延续等七种类型,却忽略社群可能是由各社群的子集合并而成的现象,如图 2 中社群 $C_{(3,4)}$ 是由社群 $C_{(2,3)}$ 和社群 $C_{(2,2)}$ 的子集合并而形成。因此,本文依据 K. Mohammadmosaferi 等定义的 12 种社群演化事件[23](见表 1)提出一种基于社群共同成员的演化事件识别算法,对所有科研社群进行演化事件识别。值得注意的是,本文将 t = 0 时刻的所有社群均视为新生社群。

表 1 科研社群演化事件定义及示例

演化事件	定义	示例
新生(form)	t 时刻出现的社群与 t-1 时刻出现的社群无共同成员(t=0 时刻所有社群均为新生社群)	$C_{(1,1)}$、$C_{(1,2)}$、$C_{(1,3)}$、$C_{(2,3)}$等
增长(grow)	t 时刻出现的社群来源于 t-1 时刻的某个社群,且该社群有新成员加入	$C_{(2,1)}$
合并(merge)	t 时刻出现的社群成员均来源于 t-1 时刻两个及两个以上的社群	$C_{(2,4)}$
合并且增长(merge and grow)	t 时刻出现的社群成员均来源于 t-1 时刻两个及两个以上的社群,并有新成员加入	$C_{(2,2)}$
部分合并(partial merge)	t 时刻出现的社群成员均来源于 t-1 时刻两个及两个以上的社群的子集	$C_{(3,4)}$
部分合并增长(partial merge and grow)	t 时刻出现的社群成员均来源于 t-1 时刻两个及两个以上的社群的子集,且该社群有新成员加入	$C_{(4,4)}$
分裂(divide)	t 时刻两个及两个以上的社群的成员均来源于 t-1 时刻的某个社群,且成员数量变少	$C_{(3,2)}$
分裂且增长(divide and grow)	t 时刻两个及两个以上的社群与 t-1 时刻的社群有共同成员,且有新成员加入	$C_{(3,3)}$
部分存活且增长(partial survive and grow)	t 时刻出现的社群只与 t-1 时刻一个社群有共同成员,且该社群有成员退出,又有新成员加入	$C_{(3,1)}$
延续(continue)	t 时刻的社群成员全部来自于 t-1 时刻的某一个社群	$C_{(4,2)}$
衰减(shrink)	t 时刻出现的社群只与 t-1 时刻一个社群有共同成员,且该社群有成员退出	$C_{(4,3)}$
消亡(dissolve)	t-1 时刻的社群与 t 时刻的社群无共同成员	$C_{(3,3)}$

3.4 科研社群演化特征分析

在科研社群演化特征分析阶段，本文主要从科研社群整体面貌、科研社群演化路径及科研社群演化特征3个方面进行探析。

在科研社群整体分析方面，本文一方面利用Cytoscape对各时期的科研合著网络进行可视化，另一方面统计各时间窗口下科研合著网络和科研社群涉及的发文量、社群数量等基础信息，并计算其节点数量、边数量及平均集聚系数等网络拓扑特征，从社会网络分析的角度理解科研合著网络和科研社群演化所展现的网络结构特征变化。

在科研社群演化路径探析方面，本文将演化事件频次、演化路径条数、起始时间等特征进行综合分析，探究科研社群在不同时期的演化关系。同时，为进一步分析科研社群的演化模式，本文统计科研社群演化树中成员数量、平均社群规模、平均发文量、演化事件频次等重要指标并进行对比分析，以期发现和揭示科研社群演化规律。

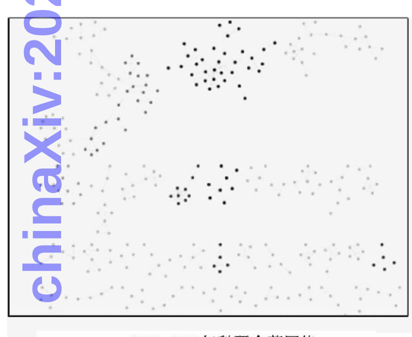
在科研社群演化特征探究方面，为反映整个领域科研社群演化特点，深挖演化事件与社群规模、发文量之间的关系，本文在进行显著性检验的基础上，对三者进行交叉分析，从而挖掘图书馆学情报学领域内科研社群演化的潜在特征。

4 结果分析

4.1 图书馆学情报学领域内科研社群整体分析

对科研社群进行整体分析可以直观揭示图书馆学情报学领域内科研社群随时间演化的整体趋势。为此，本文一方面将各时间窗口的科研合著网络进行可视化，分析网络结构随时序变化情况；另一方面对各时间窗口下科研合著网络与科研社群的基本信息、网络特征进行统计，从数值上分析其整体演化态势。

图3分别是2000－2001、2008－2009、2018－2019年图书馆学情报学领域的科研合著网络图，不同颜色的节点代表了不同的社群。从图3可看出，科研合著网络整体处于非连通状态，其中存在大量相对独立的小型科研社群，网络中最大连通图的作者节点虽然互相连通，但因合著关系的密切程度呈现出明显规模不一的社群结构。从时序角度来看，科研合著网络趋向于密集，最大连通图的规模不断增加，反映出图书馆学情报学领域呈现蓬勃发展的演变状态，科研社群之间的联系不断增强。


2000–2001年科研合著网络


2008–2009年科研合著网络

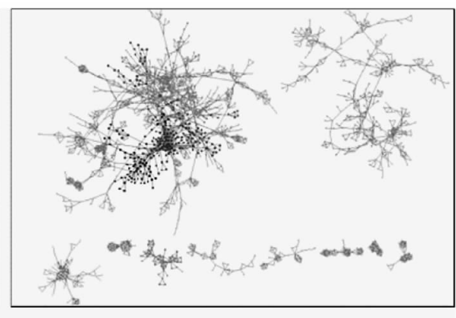

2018–2019年科研合著网络

图3 2000－2001、2008－2009、2018－2019年科研合著网络

图4为2008－2009年科研合著局部网络，各节点标签为其所在社群的编号。不难看出，在该时期图书馆学情报学领域的科研社群已初具规模，以社群编号5－61为例，该社群包含节点55个，边关系82条，社群内部联系密切且连接良好，也进一步证明Leiden算法在科研社群识别方面具有优越性。

表2为各时间窗口下科研合著网络和科研社群的基本信息。科研合著网络基本信息中，发文量随时间逐年增长，侧面说明20年来图书馆学情报学领域的研究不断深入，研究成果丰富。节点数量和边数量的变化体现科研合著网络规模的演变。节点总数的逐年上升，表明图书馆学情报学领域中合著成员在不断增加，边总数的增加在一定程度上说明图书馆学情报学领域中科研社群的合作与交流在加强，并在时间窗口9达到巅峰。平均集聚系数的逐年增强反映科研合著网络中成员之间合作频繁，合作关系不断得到稳固，“小团体”现象初显。

科研社群基本信息中，各时期科研社群数量的变化一定程度上反映了社群之间的合作演变。2015年之前（即时间窗口8）科研社群数量呈上升趋势，之后便趋于平稳，说明前期涌现的大量科研社群促进了图书馆学情报学领域的发展，后期社群之间的合作交流加强，社群合并趋势初现。科研社群与科研合著网络在发文量、节点总数、边总数等指标上也呈现出协同演

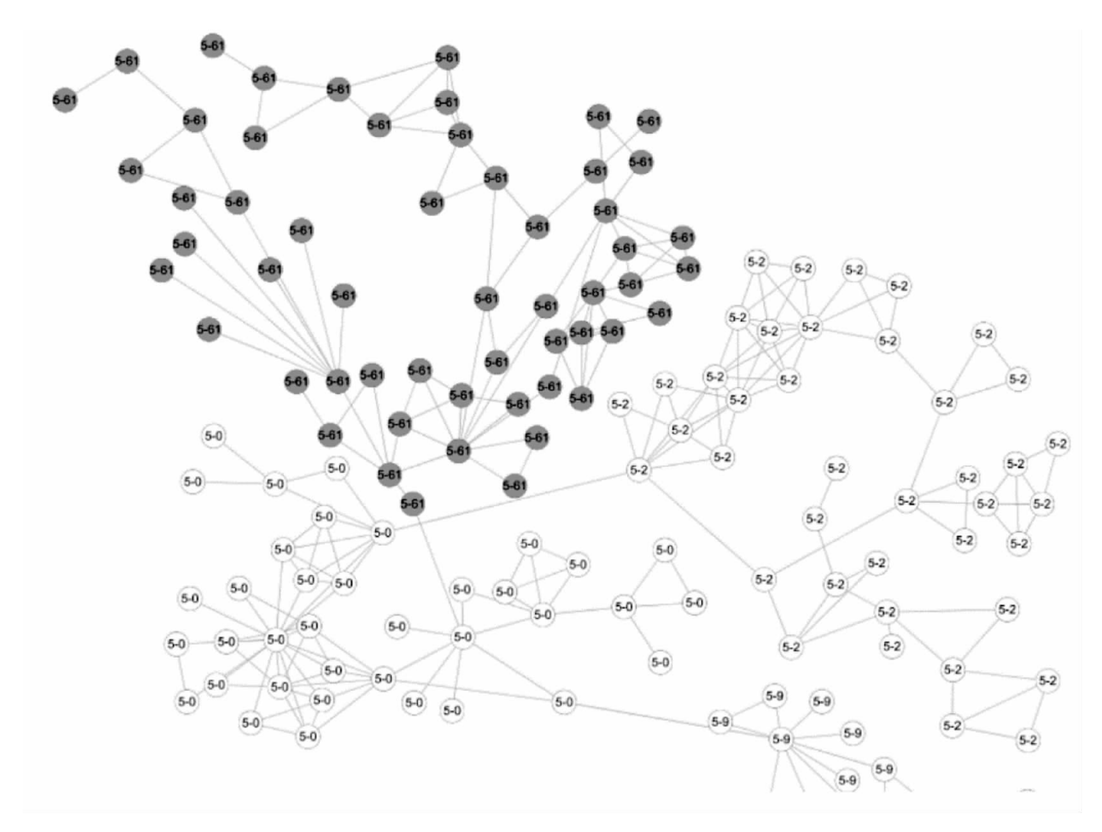


图 4　2008 – 2009 年科研合著局部网络

表 2　图书馆学情报学领域科研合著网络及社群基本信息

时间窗口	科研合著网络基本信息				科研社群基本信息			
	发文量(篇)	节点总数(个)	边总数(条)	平均聚类系数	发文量(篇)	社群数(个)	节点总数(个)	边总数(条)
1	970	1 767	791	0.146	588	101	504	582
2	1 213	2 115	1 177	0.209	780	155	770	927
3	1 631	2 972	2 002	0.251	1 033	241	1 175	1 635
4	2 413	3 994	3 075	0.304	1 743	357	1 928	2 663
5	2 601	4 513	3 677	0.333	1 847	385	2 175	3 191
6	3 175	5 130	4 900	0.341	2 292	447	2 635	4 345
7	3 455	5 732	5 469	0.364	2 578	476	3 071	4 894
8	4 162	6 313	6 643	0.385	3 224	**526**	3 560	6 061
9	4 457	6 526	**9 920**	**0.400**	3 604	486	**3 837**	**9 389**
10	4 643	**6 089**	7 158	0.360	**3 683**	448	3 413	6 618

化趋势。其中科研社群占科研合著网络发文量的比重从 60.1% 上升至 79.3%，边总数的占比从 73.5% 上升至 94.6%，说明科研社群的合作强度和科研产出能力逐年增强。

4.2　图书馆学情报学领域内科研社群演化路径分析

演化路径是揭示科研社群在不同时期演化关系的重要方法，观测演化路径上科研社群在相邻时间窗口发生的演化事件有助于从微观层面上发掘科研社群动态演化规律，而演化树则可以体现科研社群之间成员流动情况。基于匹配相邻时间窗口科研社群之间的共同成员思想，本文共构建图书馆学情报学领域内 5 796 条演化路径，其路径长度与起始点分布情况如图 5 所示，横轴为路径起始时间窗口，纵轴为路径长度。不难看出，演化路径长度大于 8 的路径较多，占总演化路径条数的 59.2%，说明图书馆学情报学领域内的科研社群具有良好的延续性。

基于 3.3 节中对科研社群演化事件的定义，本文依据演化路径数据对科研社群进行演化事件识别，得到各演化事件在时间上的分布情况，如图 6 所示，横轴代表时间窗口，纵轴代表演化事件类型。经对演化事件进行细分定义后，严格的合并、合并且增长、部分合并等演化事件出现频次较低，部分合并且增长、分裂且

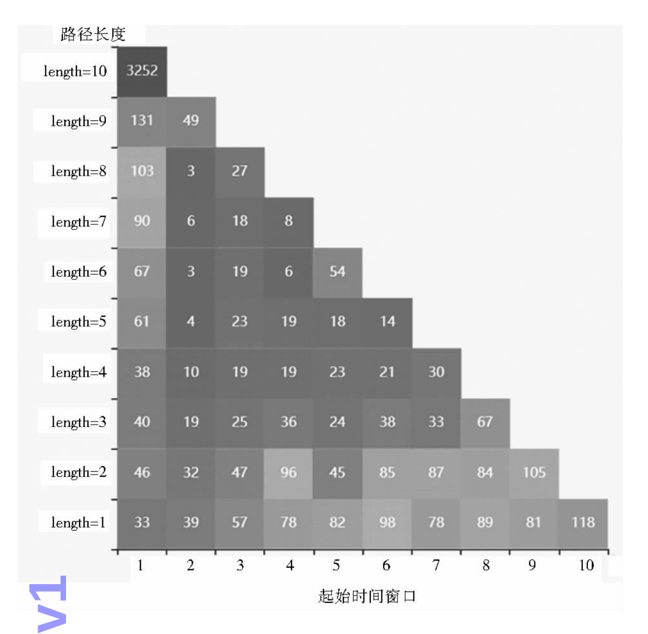


图 5　演化路径长度与起始点分布

增长、部分存活且增长等演化事件的演化频次随时间而增加，反映了该领域科研社群变化的复杂性，社群之间的合作与交流随时间推移趋向增强。新生、消亡演化事件频次较高，说明近 20 年来，图书馆学情报学领域不断有新科研社群加入的同时，也存在科研社群退出该研究领域的现象。整体而言图书馆学情报学领域中增长类演化事件频次较高，呈现出动态蓬勃发展的趋势。

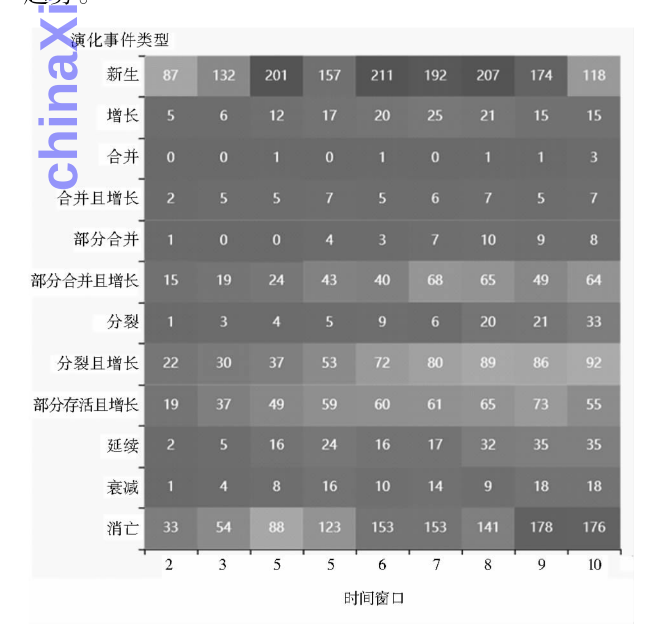


图 6　科研社群演化事件频次分布

为进一步探究科研社群的演化模式，本文对各演化树中的科研社群成员数量、平均社群规模、平均发文量、演化事件频次等指标进行统计分析，并将科研社群总数量大于 10 的演化树定义为大型演化树，反之则为小型演化树。实验最终得到 37 棵大型演化树与 1 543 棵小型演化树，其中大型演化树包含 1 363 个科研社群，小型演化树包含 3 003 个科研社群。图 7、图 8 分别为科研社群演化树中平均社群规模和平均发文量、演化事件频次等指标信息，因新生、消亡两类演化事件受演化树内社群数量影响较大，故在此不作分析。

不难看出，大型演化树虽在整个领域仅占2.34%，但其平均社群规模和平均发文量均高于小型演化树。从演化事件频次来看，大型演化树中科研社群呈现了以“部分合并且增长”和“分裂且增长”为主导的演化模式。说明在大型演化树中，合并与分裂两类演化事件交织，科研社群之间的合作交流强度较高，且不断地有新成员加入，极大地促进了科研成果的产出。

小型演化树数量众多且涉及科研社群数量大，但其平均社群规模和平均发文量处于较低水平，侧面表明小型演化树中大多为小型科研社群。从图 8 可知，小型演化树大体为以“部分合并且增长”和 “部分存活且增长”为主导的演化模式，延续、衰减等演化事件高于大型演化树，科研社群之间的合作交流程度相较于大型演化树较低。

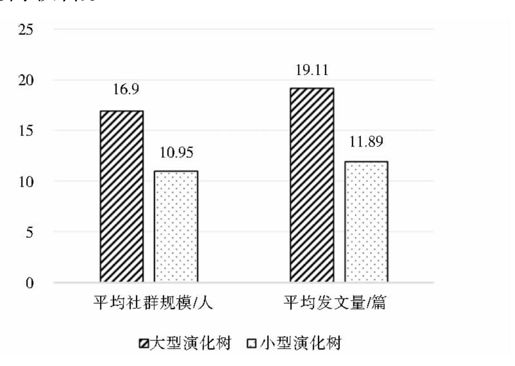


图 7　科研社群演化树基本信息

为进一步探究图书馆学情报学领域中科研社群在各时期的动态演化过程，本文采用 python 第三方可视化库 pyecharts 将该领域的科研社群演化路径以桑基图的形式展现，如图 9、图 10 所示。图中各节点代表了不同阶段的科研社群，其长度代表社群成员数量，延伸分支的宽度则代表了流入下一时期某科研社群的成员数量。

大型科研社群演化树(见表图 9)中涉及的科研社群达 197 个，包含演化路径 352 条，平均演化路径长度为6.23，说明大型科研社群演化树具有良好的延续性，继续在该领域中深入研究的社群成员比例较高。同时，大型科研社群演化树各个时期的科研社群演化事

件以

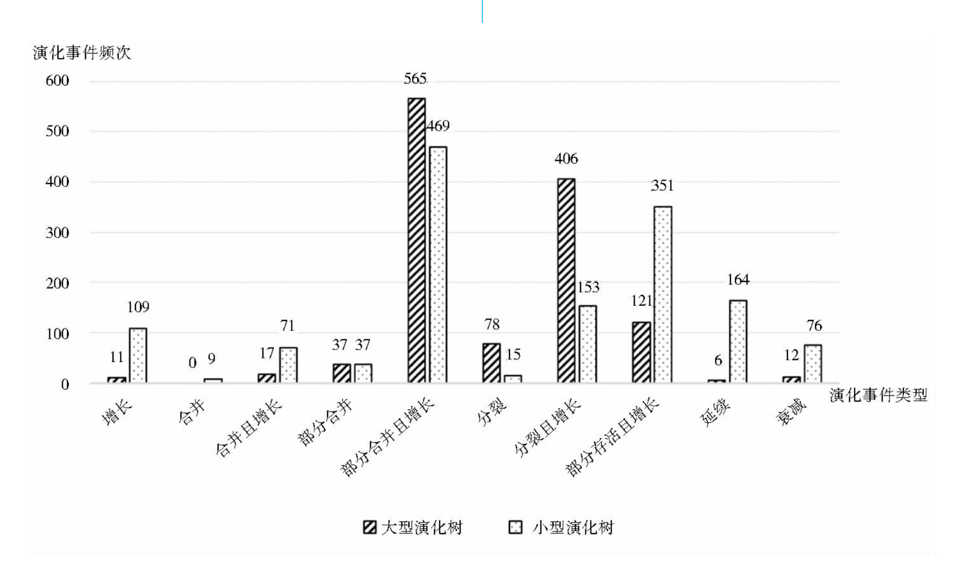


图 8 科研社群演化树中演化事件频次分布

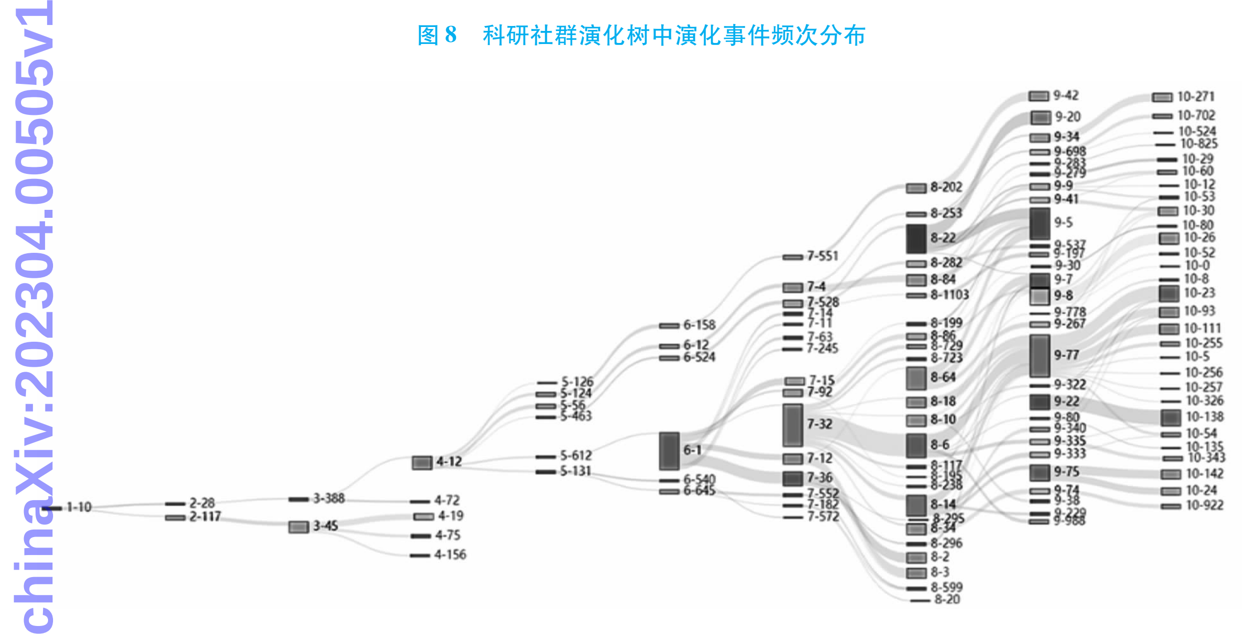


图 9 大型科研社群演化树示例

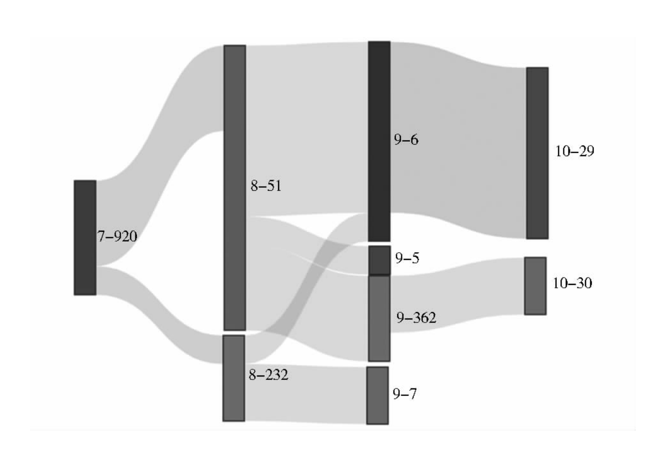


图 10 小型科研社群演化树示例

“部分合并且增长”和“分裂且增长”为主，分别占演化树中演化事件总量的 45.2%、30.5%，进一步说明大型科研社群演化树中科研社群之间交流密切，合并与分裂事件交织，演化树的规模随时间推移持续增加。相比之下小型科研社群演化树（见图 10）涉及的社群数量较少，平均路径仅为 2.67，说明该类型的演化树中退出图书馆学情报学领域研究的社群成员较多，科研社群延续性有待提升。且各时期的科研社群演化事件整体以分裂类演化事件为主导，科研社群之间的合作交流程度较低。

对科研社群的演化路径进行深入分析有助于归纳科研社群演化模式，却无法反映整个领域科研社群演化特征的问题。鉴于此，本文对于科研社群演化事件与社群规模、发文量之间的关系展开进一步探究。

4.3 图书馆学情报学领域内科研社群演化事件特征探究

李纲，唐晶，毛进，等.基于演化事件探测的学科领域科研社群演化特征研究——以图书馆学情报学为例[J].图书情报工作,2021,65(17):79－90.

演化事件检测是探究科研社群动态演化规律和趋势的重要方法，将其与科研社群规模、发文量进行综合分析可有效跟踪科研社群演化、挖掘科研社群演化特征。而变量之间差异显著才是研究的前提，为此，本文将对科研社群存在的时间窗口、社群成员数量、发文数量及演化事件进行多因素方差分析，对社群成员数量和发文数量进行相关性分析，结果如表3、表4所示：

表3　主体间效应的检验结果

源	因变量	III 型平方和	df	均方	F	Sig.
校正模型	numbers_of_members	125 095.002[a]	19	6 583.947	59.981	.000
	counts_of_papers	183 341.247[b]	19	9 649.539	48.977	.000
截距	numbers_of_members	19 324.633	1	19 324.633	176.051	.000
	counts_of_papers	16 881.058	1	16 881.058	85.681	.000
year	numbers_of_members	2 555.543	9	283.949	2.587	.006
	counts_of_papers	4 999.336	9	555.482	2.819	.003
evo_events	numbers_of_members	121 508.300	10	12 150.830	110.696	.000
	counts_of_papers	177 822.904	10	17 782.290	90.255	.000
误差	numbers_of_members	395 382.205	3602	109.767		
	counts_of_papers	709 674.971	3602	197.022		
总计	numbers_of_members	667 394.000	3622			
	counts_of_papers	1 019 124.000	3622			
校正的总计	numbers_of_members	520 477.207	3621			
	counts_of_papers	893 016.219	3621			

a. R 方 =.240(调整 R 方 =.236)；b. R 方 =.205(调整 R 方 =.201)

表4　社群成员数量与发文量相关系数

–			numbers_of_members	counts_of_papers
Spearman 的 rho	numbers_of_members	相关系数	1.000	.633 **
		Sig.(双侧)	.	.000
		N	3 622	3 622
	counts_of_papers	相关系数	.633 **	1.000
		Sig.(双侧)	.000	.
		N	3 622	3 622

** 在置信度(双测)为0.01时，相关性是显著的

从表3可以看出，对于时间窗口与演化事件这两个因素，成员数量和演化事件的P值均小于0.05，可以认定不同的时间窗口之间差异显著、演化事件之间差异显著。表4则表明成员数量和发文量的相关性为0.633，属于弱正相关关系，一定程度上说明成员数量并不是影响发文量的唯一因素，二者与演化事件的关系还需进一步深入探讨。

随后，为进一步探讨演化事件与社群成员数量、发文量的数量关系，本文将演化事件在各个时间段的平均社团规模、发文量分布情况以热力图的形式呈现，如图11、图12所示。

图11　演化事件与平均社群规模协同演化

图11中平均社群规模由该时间窗口内某演化事件中所有社群成员数量与社群数量之比所得。新生社群与消亡社群的平均社群规模较小，大部分在3－4之间，表明各个时期下新生、消亡等演化事件多发生于小型科研群体。从增长的角度来看，增长类相关演化事件的平均社群规模相较于其他演化事件较大，说明大型科研社群中更有可能发生增长类演化事件，部分原因是增长类相关演化事件与新成员加入社群息息相关。其中，部分合并与增长事件在任一时期内平均社团规模均最大，并随时间推移有规模增加的趋势，侧面说明图书馆学情报学领域内的研究人员在面对复杂研

究问题时,一方面积极引入新成员,另一方面更倾向于加强社群之间的合作与交流。从合并的角度看,合并且增长的平均社群规模较大,但严格意义上的合并、部分合并事件的平均社群规模较小,一定程度上说明图书馆学情报学领域中合并与部分合并事件大多发生于小型科研社群。分裂和衰减事件的发生即意味着有成员退出科研社群,因此平均社群规模较小。

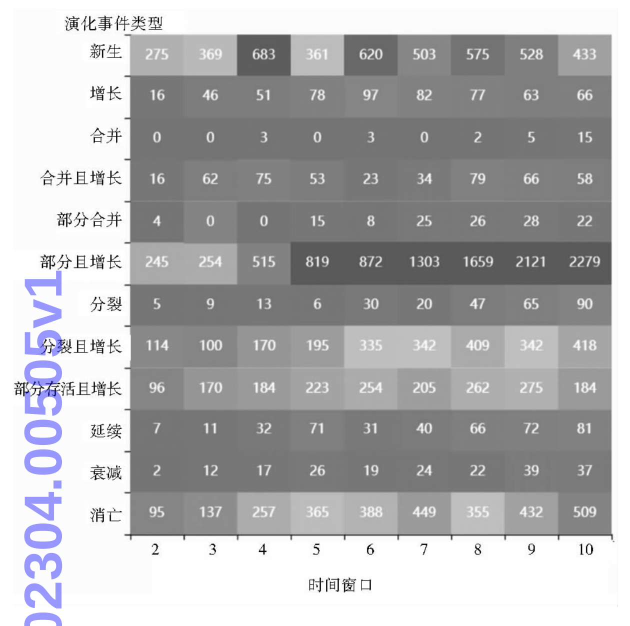


图 12　演化事件与发文量协同演化

图 12 展示了演化事件在各个时间窗口上的发文量分布情况。可以看出,增长类相关的演化事件发文量较高,且随时间呈现显著增加趋势。特别是部分合并且增长事件,该演化事件的发文量在任一时期均遥居榜首,体现了该演化事件类型下科研社群的科研能力较强。这进一步证明在图书馆学情报学领域中,社群之间的交流及新成员的加入对于科研社群的科研产出具有较大促进作用。新生和消亡演化事件所涉及的科研社群发文量也很瞩目,某种程度上和该两类演化事件包含的社群数量众多相关。另外,增长、分裂和延续事件下的科研社群虽受社群规模的影响,发文量整体偏低,但仍随时间推移呈现上升趋势,表明该几类演化事件下的社群科研水平有一定提升。

5　结论与讨论

本文以图书馆学情报学领域 2000 – 2019 年期刊文献为数据基础,采用 Leiden 算法对各时间窗口下的合著网络进行科研社群划分,并对各时期的科研社群网络特征进行分析;同时,在抽取科研社群演化路径的基础上进行社群演化事件识别,归纳出图书馆学情报学领域内两种科研社群演化模式,并通过对演化事件与社群规模、发文量进行交叉分析挖掘科研社群演化特征。综合上述对科研社群演化过程的探讨结果,本文初步得出以下结论:

(1)图书馆学情报学领域科研社群规模总体处于蓬勃发展态势,科研合著网络呈现“小团体”现象。近 20 年来,图书馆学情报学领域内科研社群的发文量、社群节点总数、边总数均呈逐期上升趋势(见表 2),各时期新生和消亡社群众多,增长类演化事件频次较高(见图 6),反映出科研社群规模呈现动态蓬勃发展态势,图书馆学情报学领域研究不断深入。同时,科研合著网络中最大连通图规模不断增大,平均集聚系数不断增加,侧面表明网络成员之间合作交流趋于频繁,形成了紧密联系的“小团体”现象。

(2)图书馆学情报学领域中科研社群呈现两种演化模式,社群之间的合作交流频繁。研究结果显示,图书馆学情报学领域大型科研社群演化树平均社群规模和平均发文量较高,呈现以“部分合并且增长”和“分裂且增长”为主导的演化模式;小型科研社群演化树平均社群规模和平均发文量较低,呈现以“部分合并且增长”和“部分存活且增长”为主导的演化模式。这意味着近 20 年来不断有新成员加入科研社群,社群之间的成员流动趋向频繁,同时也为后续进一步研究社群演化模式奠定基础。

(3)图书馆学情报学领域中科研社群演化较复杂,演化事件与社群规模、发文量存在相关性。研究表明,增长类演化事件大多发生于大型科研社群,其发文量较高,其中部分合并且增长事件无论是平均社群规模还是发文量均高居榜首,这也从侧面证明,新成员的引入以及社群之间合作交流的增强可以有效地提高科研水平、促进科研产出。新生和消亡演化事件均发生于小型科研社群,但因其在各时期出现频次较多,发文量较高。合并、部分合并、分裂、衰减等演化事件的平均社群规模较小且其发文量偏低,表明这几类演化事件更易在小型科研社群中发生,社群的规模也会在一定程度上影响发文量。这说明不同的演化事件呈现的特征也会有一定差异,进一步反映出图书馆学情报学领域中科研社群演化的复杂性。

本文提出了一种探究学科领域中科研社群演化特征的方法,采用 Leiden 算法识别、提取图书馆学情报学领域中的科研社群,并对其进行动态社群演化跟踪与分析。研究工作一方面从时序上对科研社群网络演化特征进行整体分析,另一方面通过抽取科研社群演化

路径和识别演化事件,对科研社群演化模式进行归纳,最后对科研社群的演化事件特征进行探究。本研究丰富了科研社群演化的相关理论与方法,进一步揭示图书馆学情报学领域内科研社群演化的模式与特征,有助于促进科研社群的发现、组建与管理。本文还存在需要完善的地方,如未考虑科研人员在非连续时间窗口发文,未将主题融入科研社群演化特征探究等。在未来研究工作中,将进一步探讨主题与科研社群的协同演化过程,以期更全面地对领域中科研社群的演化模式与特征进行探索与揭示。

参考文献:

[1] 任妮,周建农.合著网络加权模式下科研团队的发现与评价研究[J].现代图书情报技术,2015(9):68-75.

[2] EVANS T S, LAMBIOTTE R, PANZARASA P. Community structure and patterns of scientific collaboration in business and management[J]. Scientometrics, 2011, 89(1):381-396.

[3] MAO J, CAO Y, LU K, et al. Topic scientific community in science: a combined perspective of scientific collaboration and topics[J]. Scientometrics, 2017, 112(4):851-875.

[4] LIU X, BOLLEN J, NELSON M L, et al. Co-authorship networks in the digital library research community[J]. Information processing & management, 2005,41(6):1462-1480.

[5] 李亮,朱庆华.社会网络分析方法在合著分析中的实证研究[J].情报科学,2008,26(4):549-555.

[6] MASMOUDI A, MEZGHANI E, BELLAAJ H, et al. A web-based knowledge management system for scientific research team[C]//2017 IEEE 26th international conference on enabling technologies: infrastructure for collaborative enterprises. Poland:IEEE, 2017.

[7] 陈文杰.融合节点主题特征的社团发现研究[J/OL].情报理论与实践:1-10[2021-01-22]. http://kns.cnki.net/kcms/detail/11.1762.g3.20201223.0841.002.html.

[8] JUNG S, YOON W C. An alternative topic model based on common interest authors for topic evolution analysis[J]. Journal of informetrics, 2020, 14(3):101040.

[9] ROSVALL M, BERGSTROM C T. Maps of random walks on complex networks reveal community structure[J]. Proceedings of the national academy of sciences, 2008, 105(4):1118-1123.

[10] RAGHAVAN U N, RÉKA ALBERT, KUMARA S. Near linear time algorithm to detect community structures in large-scale networks[J]. Physical review E, 2007, 76(3 Pt 2):036106.

[11] BLONDEL V D, GUILLAUME J L, LAMBIOTTE R, et al. Fast unfolding of communities in large networks[J]. Journal of statistical mechanics theory & experiment, 2008, 10: P10008.

[12] 孙扬.社交网络中社区演化事件检测研究[D].上海:上海交通大学,2016.

[13] 余厚强,白宽,邹本涛,等.人工智能领域科研团队识别与领军团队提取[J].图书情报工作,2020,64(20):4-13.

[14] 邹本涛,王曰芬,余厚强.人工智能领域高产科研团队的演化研究[J].图书情报工作,2020,64(20):23-33.

[15] 江文华,徐健,李纲,等.基于研究兴趣相似性网络的我国图书馆学研究社群分析[J].现代情报,2019,39(9):21-27.

[16] TRAAG V, WALTMAN L, VAN ECK N J. From louvain to leiden: guaranteeing well-connected communities[J]. Scientific reports,2019,9(1):5233

[17] 韩童茜,王立梅,许鑫.长三角城市群科研合作网络演化研究——基于SCIE和SSCI论文的实证分析[J].情报理论与实践,2020,43(10):151-156.

[18] 李纲,李春雅,李翔.基于社会网络分析的科研团队发现研究[J].图书情报工作,2014,58(7):63-70,82.

[19] 王曰芬,李冬琼,余厚强.生命周期阶段中的科学合作网络演化及高影响力学者成长特征研究[J].情报学报,2018,37(2):121-131.

[20] ASUR S, PARTHASARATHY S, UCAR D. An event-based framework for characterizing the evolutionary behavior of interaction graphs[J]. ACM transactions on knowledge discovery from data, 2009, 3(4):1-36.

[21] TAKAFFOLI M, FAGNAN J, SANGI F, et al. Tracking changes in dynamic information networks[C]//International conference on computational aspects of social networks. Spain:IEEE, 2011.

[22] PALLA G, BARABÁSI, ALBERT-LÁSZLÓ, VICSEK, TAMÁS. Quantifying social group evolution[J]. Nature, 2007(446):664-667.

[23] MOHAMMADMOSAFERI K K, NADERI H. Evolution of communities in dynamic social networks: an efficient map-based approach[J]. Expert systems with applications, 2020(147):113221.

[24] 汤强.科研网络的社区发现及演化特征研究[D].西安:西安电子科技大学,2015.

[25] BAE S H, D HALPERIN, WEST J D, et al. Scalable and efficient flow-based community detection for large-scale graph analysis[J]. ACM transactions on knowledge discovery from data, 2017, 11(3):1-30.

[26] 胡昌平,陈果.层次视角下概念知识网络的三元关系形态研究[J].图书情报工作,2014,58(4):11-16.

[27] SHANNON, P. Cytoscape: a software environment for integrated models of biomolecular interaction networks[J]. Genome research, 2003, 13(11):2498-2504.

[28] 徐兵,赵亚伟,徐杨远翔.基于关联群演化相似度的社团追踪算法[J].复杂系统与复杂性科学,2019,16(1):14-25.

作者贡献说明:

李纲:确定论文思路,提出修改意见;

唐晶:设计论文思路,处理与分析数据,论文撰写;

毛进:提出问题,审核与修改论文;

田云裴:处理与分析数据;

张斌:分析数据,修改论文。

Research on the Evolution Characteristics of Scientific Research Communities in Subject Fields Based on Evolutionary Event Detection-An Example of LIS

Li Gang[1] Tang Jing[1] Mao Jin[1] Tian Yunpei[1] Zhang Bin[2]
[1] Center for the Studies of Information Resources, Wuhan University, Wuhan 430072
[2] School of Information Management, Nanjing University, Nanjing 210023

Abstract: [**Purpose/significance**] Scientific research communities are important knowledge groups in contemporary science. Studying the evolutionary characteristics of scientific research communities is of great significance for exploring the law of field development and promoting knowledge innovation. [**Method/process**] This article took the field of Library and Information Science (LIS) as an example. From the perspective of evolutionary event detection, this paper used the Leiden algorithm to detect scientific research communities, and constructed their evolution paths and evolution trees. On this basis, this paper identified the evolution events of scientific research communities, and revealed the evolution modes and evolution characteristics of scientific research communities from three aspects: the overall analysis of the evolution, the evolution paths and the characteristics of the evolution trees, and the statistical characteristics of group evolution events. [**Result/conclusion**] The research shows that the scale of scientific research communities is developing vigorously, and the evolution trees of scientific research communities present two evolution modes. Most of the evolutionary events of growth type occurred in large communities with a relatively high volume of posts, while both 'form' and 'dissovle' evolution events occurred in small communities with a relatively high volume of posts. The average community size of evolutionary events such as 'merge', 'partial merge', 'split', and 'shrink' is small, and the volume of publications is low. These characteristics further prove that the cooperation and exchanges between scientific research communities tend to be frequent, and the evolution of scientific research communities has become increasingly complex.

Keywords: scientific research community Leiden algorithm evolutionary events evolutionary characteristics